ehp ENVIRONMENTAL HEALTH PERSPECTIVES
http://www.ehponline.org

Mechanism Profiling of Hepatotoxicity Caused by Oxidative Stress Using the Antioxidant Response Element Reporter Gene Assay Models and Big Data

Marlene Thai Kim, Ruili Huang, Alexander Sedykh, Wenyi Wang, Menghang Xia, and Hao Zhu

http://dx.doi.org/10.1289/ehp.1509763

Received: 27 January 2015
Accepted: 16 September 2015
Advance Publication: 18 September 2015

Environ Health Perspect DOI: 10.1289/ehp.1509763
Advance Publication: Not Copyedited

Mechanism Profiling of Hepatotoxicity Caused by Oxidative Stress Using the Antioxidant Response Element Reporter Gene Assay Models and Big Data

Marlene Thai Kim,[1,2] Ruili Huang,[3] Alexander Sedykh,[2,4] Wenyi Wang,[2] Menghang Xia,[3] and Hao Zhu[1,2]

[1]Department of Chemistry, Rutgers University, Camden, New Jersey, USA; [2]Center for Computational and Integrative Biology, Rutgers University, Camden, New Jersey, USA; [3]National Center for Advancing Translational Sciences, National Institutes of Health, Bethesda, Maryland, USA; [4]Multicase Inc., Beachwood, Ohio, USA

Address correspondence to Hao Zhu, 315 Penn St., Rutgers University, Camden, NJ 08102 USA. Telephone: (856) 225-6781. E-mail: hao.zhu99@rutgers.edu

Running title: Toxicant profiling using big data sources

Acknowledgments: Research reported in this publication was supported, in part, by the National Institute of Environmental Health Sciences of the National Institutes of Health under Award Number R15ES023148 and the Colgate-Palmolive Grant for Alternative Research. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health. Alexander Sedykh is employed by Multicase Inc., Beachwood, Ohio.

Competing financial interests: All authors declare they have no actual or potential competing financial interests.

Abstract

Background: Hepatotoxicity accounts for a substantial number of drugs withdrawn from the market. Traditional animal models used to detect hepatotoxicity are expensive and time consuming. Alternative *in vitro* methods, especially cell-based High-Throughput Screening (HTS) studies, have provided the research community with a large amount of data from toxicity assays. Among the various assays used to screen potential toxicants is the Antioxidant Response Element *beta* lactamase reporter gene assay (ARE-*bla*), which identifies chemicals that have the potential to induce oxidative stress and was used to test > 10,000 compounds from the Tox21 program.

Objective: The ARE-*bla* computational model and HTS data from a big data source (PubChem) was used to profile environmental and pharmaceutical compounds with hepatotoxic data.

Methods: Quantitative Structure-Activity Relationship models were developed based on ARE-*bla* data. The models predicted the potential oxidative stress response for known liver toxicants when there was no ARE-*bla* data available. Liver toxicants were used as probe compounds to search PubChem Bioassay and generate a response profile, which contained thousands of bioassays (> 10 million data points). By ranking the *In Vitro-In Vivo* Correlations (IVIVC), the most relevant bioassay(s) related to hepatotoxicity were identified.

Results: The liver toxicants profile contained the ARE-*bla* and relevant PubChem assays. Potential toxicophores for well-known toxicants were created by identifying chemical features that existed only in compounds with high IVIVC.

Conclusion: Profiling the chemical IVIVCs created an opportunity to fully explore the source-to-outcome continuum of modern experimental toxicology using cheminformatics approaches and big data sources.

Introduction

Traditional animal models used to evaluate hepatotoxicity are expensive and time consuming (Hartung 2009). *In vitro* assays are used as an alternative to better understand hepatotoxicity (Adler et al. 2011; Zhu et al. 2014a). However, endeavors to correlate *in vitro* and *in vivo* hepatotoxicity (Moeller 2010) have not successfully replaced *in vivo* hepatotoxicity models (Ekins 2014; MacDonald and Robertson 2009).

There is an unmet need to develop predictive assays for hepatotoxicity (Chen et al. 2014). As an alternative, High-Throughput Screening (HTS) approaches are used to screen large chemical libraries (> 50,000 compounds) to elucidate toxic mechanisms and prioritize candidates for further animal tests (Zhu et al. 2014b). This led to the rapid generation of bioassay data. PubChem, the leading public bioassay data repository, contains > 50 million compounds and > 700,000 assays (Wang et al. 2014). This amount of "big data" is difficult to process and analyze using standard data processing tools.

Another issue with using HTS for toxicological studies is that it tests compounds at one concentration, which may not reveal its toxic effects. This was addressed by the US Tox21 inter-agency collaboration (Attene-Ramos et al. 2013; Collins et al. 2008; Committee on Toxicity Testing and Assessment of Environmental Agents 2007; Dix et al. 2007). Based on their guidelines, the National Institutes of Health Chemical Genomics Center (NCGC), now part of the National Center for Advancing Translational Sciences (NCATS), developed Quantitative High-Throughput Screening (qHTS) (Inglese et al. 2006). A qHTS experiment tests > 100,000 compounds at 15 different concentrations in triplicate within a week (Attene-Ramos et al. 2013). This approach is more rational than single-dose HTS, because it simulates dose-dependent

animal toxicity effects (Eaton and Gilbert 2010). These results are available online (http://www.ncbi.nlm.nih.gov/pcassay?term=tox21, accessed January 19, 2015).

The Antioxidant Response Element (ARE) pathway plays a major role in regulating and alleviating oxidative stress (Ma 2013), which after long-term exposure causes many pathophysiological conditions, including cancers and hepatotoxicity (Hybertson et al. 2011; Shuhendler et al. 2014). Briefly, the ARE pathway is regulated by Kelch-like ECH-associating protein 1 (Keap1) and nuclear factor erythroid 2-related factor 2 (Nrf2). Keap1 contains cysteine residues that interact with reactive oxygen species (ROS) and electrophilic fragments that can trigger the dissociation of the Keap1-Nrf2 complex (Zhang and Hannink 2003). Then, Nrf2 translocates into the nucleus (Kensler et al. 2007), binds to the ARE (Itoh et al. 1997), and regulates the transcription of the antioxidative enzymes (Venugopal and Jaiswal 1998). Hindering transcription can lead to the accumulation of ROS, oxidative stress, and liver toxicity (Shuhendler et al. 2014). The qHTS ARE *beta* lactamase reporter gene assay (ARE-*bla*) can detect compounds that activate the ARE pathway and induce oxidative stress (Attene-Ramos et al. 2013; Shukla et al. 2012; Simmons et al. 2011). However, this assay alone is not sufficient for accessing animal toxicity. The correlations between the ARE pathway and animal toxicity (*i.e.,* hepatotoxicity) are not well understood.

Even with all the data from HTS and/or qHTS studies, the relationship between *in vitro* and *in vivo* toxicity is still unclear (Low et al. 2011; O'Brien et al. 2006). In this study, this challenge was addressed by developing chemical *in vitro-in vivo* correlations (IVIVC) between ARE pathway activation and hepatotoxicity (*i.e.,* liver damage). An in-house automated profiling tool and cheminformatics approaches used qHTS ARE-*bla* and liver toxicity data to retrieve relevant assays, from PubChem, and revealed liver toxicity targets. Analyzing chemical

fragments of liver toxicants revealed potential toxicophores (toxic chemical features) with clear IVIVC for a subset of compounds. Our study suggests that the use of assays as an alternative model for toxicity is feasible based on the chemical IVIVC identified from a big data source.

Methods

Databases. **qHTS ARE-*bla* Dataset.** The initial concentration-response profiles for the Tox21 10K collection tested in the qHTS ARE-*bla* tests were conducted at the NCATS (Attene-Ramos et al. 2013; Shukla et al. 2012). The Tox21 10K chemical library (http://www.epa.gov/ncct/dsstox/sdf_tox21s.html, accessed October 2, 2012) consists of compounds procured from commercial sources by the Environmental Protection Agency (EPA), National Toxicology Program (NTP), and NCGC (Huang et al. 2011), for a total of ~10,500 plated compound solutions consisting of 8,311 unique chemical substances including pesticides, industrial, food-use, and drugs. The qHTS ARE-*bla* datasets can also be downloaded from PubChem using Bioassay Accession Identifiers (AID) 743219 and 651741. PubChem is a public repository for chemical structures and their biological properties (Wang et al. 2014). Bioactivity data in PubChem are contributed by hundreds of institutes, research laboratories, and specifically those screening centers under the NIH Molecular Libraries Program (MLP) and the Tox21 program. Descriptions of the individual datasets are listed in Table 1.

The concentration-responses were normalized, range-scaled to [0, 100], and converted into curve fingerprints (Sedykh et al. 2011) using an in-house program. The source code can be downloaded from GitHub (https://github.com/sedykh/curvep). Each curve fingerprint was summed into one value termed "CurveP." CurveP represents the overall signal of the compound from its qHTS concentration-response curve that was noise filtered (*e.g.,* CurveP = 0 means no significant signals observed). Three criteria were used to classify each compound with regard to

activity: 1) CurveP, 2) maximum concentration-response, and 3) number of concentration-responses ≥ 20. The latter two describe the consistency in the concentration-responses. The scheme is detailed in Table 2. For example, a compound was classified as active if CurveP was > 0 and more than one concentration-response ≥ 20. Lastly, since all compounds were tested multiple times and all data were available, activities of each compound were averaged before classification.

***In Vivo* Hepatotoxicity Dataset.** A liver damage dataset compiled by the Food and Drug Administration (FDA) Center for Drug Evaluation and Research (Zhu and Kruhlak 2014) and Multicase Inc., contained 1,314 compounds (661 toxic and 653 non-toxic)**.**

Chemical Structure Curation. The structures of all compounds used in this study were curated to remove errors and standardized to a uniform representation. Konstanz Information Miner (KNIME) version 2.9.2 matched all compound names and PubChem Compound Accession Identifiers (CID) with its appropriate Simplified Molecular-Input Line-Entry System (SMILES) from PubChem. The in-house descriptor generators could not process large molecules (molecular weight > 2000 g/mol) and compounds without chemical structures. These compounds were removed. ChemAxon Standardizer and Structure Checker version 6.2.2 and CASE Ultra version 1.5.0.1 curated, standardized, and converted all the chemical structures into 2-D SMILES. Stereoisomers were considered as one compound. Metalorganics were removed and all salts were neutralized, because the descriptor generator cannot process them. Mixtures were manually evaluated and the major component was kept.

Measures of Quality and Reliability. To systematically evaluate the quality and reliability of the Quantitative Structure-Activity Relationship (QSAR) models and IVIVCs developed in this study, we calculated the sensitivity and specificity of each assay relative to *in vivo* animal

toxicity data, and derived the correct classification rate (CCR) where CCR = [(sensitivity + specificity) / 2] × 100 (Daniel 2009; Kim et al. 2014). In addition, we calculated the likelihood parameter (*L*) as an indication of the likelihood that active responses in a bioassay correlated with *in vivo* toxicity outcomes, where L = sensitivity × [(false positives + true positives)/ (false positives + 1)] (Zhang et al. 2014). The statistical significance of the IVIVCs were determined using Chi square (X^2) tests comparing the *in vitro* assay predictions to expectations based on *in vivo* toxicity data, under the null hypothesis of no association between the two data sources (Daniel 2009).

Workflow for Profiling the Mechanisms of Liver Toxicants. The chemical IVIVC between qHTS ARE-*bla* perturbation or relevant PubChem assays and liver damage was evaluated. The profiling workflow has three major stages (Figure 1): 1) automated biological response profiling, 2) QSAR modeling of qHTS ARE-*bla* activation, 3) chemical IVIVC evaluation.

Automated Biological Response Profiling. The biological response profile was constructed from PubChem Bioassay data (http://www.ncbi.nlm.nih.gov/pcassay/, accessed February 27, 2014) with an in-house automated profiling tool (Zhang et al. 2014), which resulted in two profile groups. One group was related to qHTS ARE-*bla* activation and the second was related to liver damage. The correlations between all bioassays (> 2,000) and ARE-*bla* and liver damage were calculated (sensitivity, specificity, CCR, and L). Only bioassays that fit the following criteria were considered for the final biological response profile: 1) appeared in both profile groups; 2) contained > 10 active responses that matched the inputted data; 3) correlation was better than random (CCR > 0.5 and L ≥ 1); and 4) is an *in vitro* assay. Lastly, bioassays were selected for further analysis if there was literature evidence that showed these assays were used to study oxidative stress and/or liver damage.

It was hypothesized that compounds that were active in multiple assays, but were not pan assay interference compounds (Baell and Holloway 2010) (*i.e.,* compounds that show false positive results in many assays due to assay technology specific artifacts), were more likely to be toxic compounds. Using the responses from the selected assays, the Rate of Actives (RA) was calculated to represent all the bioassay responses for each compound:

$$Rate\ of\ actives = \frac{A}{A+I}, \quad (1)$$

where A is the number of active responses and I is the number of inactive responses for a compound. The RA parameter was designed for this big data research since missing data can occur in the response profiles for target compounds. For example, if four assays were identified and a compound tested in all four assays was active in one assay, and negative in the other three assays, it would have a RA = 0.25. However, if another compound was active in one assay, negative in two assays, and has no data or an inconclusive result for the fourth assay, it would have a RA = 0.33. Thus, potential bias due to missing assay data was reduced. An arbitrary RA threshold was used to distinguish toxic from non-toxic compounds (RA > 0.25 as toxic, RA ≤ 0.25 as non-toxic). The RA values were used to determine the IVIVC between liver damage and the assays. To measure the quality and reliability, each RA value was classified as true positive (TP), true negative (TN), false positive (FP), or false negative (FN) for a X^2 test ($\alpha = 0.05$).

QSAR Modeling of the ARE-bla Pathway. The qHTS ARE-*bla* datasets were used to develop qHTS ARE-*bla* combinatorial QSAR models. 2-D chemical descriptors for each compound were generated using Molecular Operating Environment (MOE) version 2011.10 and Dragon 6 version 6.0. All descriptors were normalized and range scaled to [0, 1]. 186 MOE and 2,629 Dragon descriptors were used to model qHTS ARE-*bla* activation.

The qHTS ARE-*bla* dataset was down-sampled using a chemical similarity search approach to balance the ratio of active and inactive compounds selected for modeling (Sedykh et al. 2011; Willett et al. 1998). This prevents the development of biased models. Active and inactive compounds from the Tox21 phase II dataset were selected to create the modeling set, since it was much larger than the Tox21 phase I dataset (Golbraikh et al. 2003; Tice et al. 2013). Using all 186 MOE descriptors, a principal component analysis was performed. Individual models were developed using the combination of MOE or Dragon descriptors and with either Random Forest (RF) (Breiman 2001), Support Vector Machine (SVM) (Vapnik 2000), or *k*-Nearest Neighbor (*k*-NN) (Zheng and Tropsha 2000) algorithms. Six different combinations of descriptors and algorithms were used for modeling: MOE-RF, MOE-SVM, MOE-*k*-NN, Dragon-RF, Dragon-SVM, and Dragon-*k*-NN. Modeling results were averaged into a consensus model. Models were validated using 5-fold external cross-validation (80/20% split). Additional details about QSAR modeling and validation approaches can be found elsewhere (Golbraikh et al. 2003; Kim et al. 2014; Tropsha and Golbraikh 2007).

Since prediction values ranged from [0,1], two Consensus Prediction Thresholds (CPT) (Kim et al. 2014) were defined to classify compounds as active or inactive: CPT-1($\geq$0.5 as actives and < 0.5 as inactives) and CPT-2 ($\geq$ 0.8 as actives and $\leq$ 0.3 as inactives). Predictions between CPT-2 thresholds (< 0.8 and > 0.3) were inconclusives. An Applicability Domain (AD) determined whether the external compounds were structurally dissimilar to the modeling set compounds or not (Tropsha and Golbraikh 2007). Predictions of compounds outside the AD were considered unreliable. Therefore, the coverage (fraction of compounds that are within the AD) was calculated when applying AD to the predictions.

Chemical IVIVC Evaluation. Potential toxicophores, chemical fragments with significant IVIVC, were identified by inputting compounds active in the qHTS ARE-*bla* and liver damage datasets into CASE Ultra and ChemoTyper version 1.0. The substructure search tool in KNIME searched the qHTS ARE-*bla* and liver damage datasets for compounds containing the potential toxicophores. The qHTS ARE-*bla* combinatorial QSAR models predicted compounds from the liver damage dataset that have not been tested in the qHTS ARE-*bla* assay. The predictions were classified as TP, TN, FP, or FN to evaluate the chemical IVIVC for each subset of compounds with the potential toxicophores. The chemical IVIVC results were indicated using sensitivity, specificity, CCR, X^2 ($\alpha = 0.05$) (Daniel 2009).

Results

Overview of qHTS ARE-bla Dataset. The original qHTS ARE-bla data contained two datasets (Tox21 phase I and phase II). After combining, curating, and standardizing the chemical structures and activities, 6,767 unique compounds (919 actives, 748 potential actives, 760 inconclusives, and 4,340 inactives) remained. Potential active and inconclusive compounds were excluded from further analyses. The remaining Phase I dataset consists of 1,474 unique compounds (341 actives and 1,133 inactives) and Phase II dataset consists of 5,134 unique compounds (878 actives and 4,256 inactives).

qHTS ARE-bla Combinatorial QSAR Models. Six individual and one consensus qHTS ARE-*bla* QSAR models were developed for the modeling set (7 models total). The down-sampled modeling set contained 1,550 (750 actives and 800 inactives) unique compounds. Compounds left out of the modeling sets were placed into external validation sets. The chemical space, in a 3-D plot, covered by the modeling set versus its left out compounds and the liver damage dataset are shown in Figure 2A and 2B, respectively. External validation sets I [from Tox21 phase I]

and II [for Tox21 phase II] contained 1,148 (175 active and 973 inactive) and 3,584 (128 active and 3,456 inactive) compounds, respectively. The predictions of these QSAR models for new compounds represent the potential effect of these chemicals (either activation or no effect) in the qHTS ARE-*bla*.

The performance of the qHTS ARE-*bla* combinatorial QSAR consensus model in the 5-fold cross validation and against the external validation sets, with an AD for CPTs 1 and 2 are shown in Table 3. The consensus modeling set showed good performance in the 5-fold cross validation (sensitivity = 75-76%, specificity = 71-92%, and CCR = 74-84%). The performance of the consensus model against external validation sets I and II without AD was satisfactory (sensitivity = 68-93%, specificity = 72-99%, and CCR = 77-92%). Using an AD, the external validation sets still resulted in acceptable performance (sensitivity = 62-90%, specificity = 78-99%, CCR = 79-93%, coverage = 34-77%). The individual models showed acceptable performance in the 5-fold cross validation (sensitivity = 68-77%, specificity = 58-73%, and CCR = 67-73%) (Supplemental Material, Figure S1). Overall, the consensus prediction results are comparable to the results of the best individual model which is Dragon-RF (sensitivity = 74%, specificity = 73%, CCR = 73%) (Supplemental Material, Figure S1).

Liver Toxicants Profile and Its IVIVCs. The goal of the automatic data mining and extraction tool used in this study is to reduce the big data pool to a much smaller size, which can be curated manually by experts. The profiling tool identified 2,978 assays (available upon request from the corresponding author) relevant to qHTS ARE-*bla* activation and/or liver damage, 958 of which existed in both profiles. Automated data extraction identified 20 PubChem assays based on the first three criteria for assay selection (appeared in both profile groups, contained > 10 active responses that matched the inputted data, CCR > 0.5 and L ≥ 1). The assays are listed in

Supplemental Material, Table S1. However, automatic methods cannot detect the detailed characteristics of an assay and distinguish the difference between *in vitro* and *in vivo* assays. The 20 assays identified by the initial automated screening procedure were manually reviewed to confirm that they met the *in vitro* selection criterion. For example, AID 1199, was identified as an *in vivo* assay. It did not fit the "*in vitro* assay" criterion and was removed. A total of eight non-*in vitro* assays were removed in this step and there were 12 *in vitro* assays left. Through the literature search, there is no information to support the relevance of six assays (AIDs 121, 123, 589, 590, 2330, and 720532) to either liver damage or oxidative stress. Six assays remained and two of them had redundant activities. For example, AIDs 686978 and 686979 refer to the qHTS human tyrosyl-DNA phosphodiesterase 1 (TDP1) assay tested under two different conditions, and the activities for most of the compounds were the same. AID 686978 was selected since the condition was performed in absence of the topoisomerase I poison camptothecin, which was more suitable for this study. AIDs 743065 and743067 refer to the qHTS assay to identify small molecule antagonists of the thyroid receptor (TR) signaling pathway. AID 743067 was selected because it was a summary assay (included both primary and cell viability counter screen results). After removing the redundant assays and evaluating the remaining assays by their mechanisms, four PubChem assays remained: AID 686978 qHTS for inhibitors of TDP1, AID 743067 qHTS assay to identify small molecule antagonists of the TR signaling pathway, AID 743140 qHTS assay to identify small molecule agonists of the peroxisome proliferator-activated receptor gamma (PPARg) signaling pathway, and AID 743202 which was the qHTS ARE-*bla* assay used in the QSAR models above. These assays are relevant to ARE perturbation and liver damage according to literature (Fielden et al. 2007; Königer et al. 2014; Malik and Hodgson 2002; Mantena et al. 2008) and were combined to create the biological response profile (Figure 3A).

Although these top four assays met the selection criteria, the individual assay predictions were not significantly associated with *in vivo* liver damage (X^2 *p*-values for the independence of assays and *in vivo* liver damage = 0.24-0.59). However, combining these four assays and defining toxicity as RA > 0.25 resulted in a statistically significant association (X^2 *p*-value = 2.92×10^{-4}). The biological profile shows the responses for 953 compounds from the liver damage dataset against the top four assays and their combined responses, using threshold RA > 0.25 (Figure 3A). 361 liver damage compounds are not shown, because there was no bioassay data available for them.

The qHTS ARE-*bla* dataset used in this study contains > 6,000 compounds, but does not cover all the compounds in the liver damage dataset. Therefore, qHTS ARE-*bla* combinatorial QSAR model was used to predict the activity of compounds that were not tested in the qHTS ARE-*bla* study. It is important to mention that the liver damage dataset consists of mostly drug-like compounds that were outside of the AD of the QSAR models. In previous studies, QSAR models normally cannot predict the compounds out of AD as accurately as the compounds within AD (Tropsha and Golbraikh 2007). As shown in the principal component analysis (Figure 2B) and according to the AD analysis, most of the liver damage dataset compounds either share the same chemical space as the actives in the modeling set or are out of AD, meaning they are likely to be predicted as active by the QSAR models. This resulted in the increase of false positives in the later IVIVC analysis, which provides a hint that extra experimental ARE data are still needed for the drug-like compounds of interest in the future study.

Using CASE Ultra and ChemoTyper, two subsets of compounds were identified. Subsets contained a chemical fragment that showed a statistically significant IVIVC between ARE-*bla* activation and liver damage in the X^2 test with *p*-values of 0.01 and are referred to as potential

toxicophores A and B (Figure 3B), respectively. There are more true positives than false positives. Therefore, the active responses in this assay are potential signals of liver damage for the compounds that contain the potential toxicophores.

Furthermore, the qHTS ARE-*bla* combinatorial QSAR models were used to predict liver damage dataset compounds without experimental qHTS ARE-*bla* perturbation results. Figure 3B shows the IVIVC (TP, TN, FP, and FN) between the qHTS ARE-*bla* activation and liver damage, for compounds with potential toxicophores A and B, using experimental ARE-*bla* data and QSAR predictions. When using only QSAR results, the IVIVC was not statistically significant (X^2 p-value = 0.41) for both potential toxicophores. This is due to structural differences between the drugs in the liver damage dataset and the compounds in the Tox21 dataset, used to develop the qHTS ARE-*bla* combinatorial QSAR model, as described above. The result suggests the limitation of applying QSAR models to predict new compounds that are out of AD.

Discussion

ARE pathway perturbation is an important mechanism for alleviating and preventing oxidative stress (Ma 2013). In this study, qHTS ARE-*bla* data and the resulting QSAR models were used to study the relationship between oxidative stress and liver damage. When qHTS ARE-*bla* data for a compound was not available, the combinatorial QSAR models were used to fill-in the empty entries. This technique can be adapted to populate response profiles for other assays.

The workflow created in this study used data from PubChem, a publicly available big data source, to create and populate a bioassay response profile and revealed the relationship between oxidative stress and liver damage (Figure 1). Furthermore, the workflow in this study

can be adapted to develop adverse outcome pathways (AOP) (Ankley et al. 2010). Our study identified a combination of molecular initiating events (MIE) (Allen et al. 2014) between some drugs and biomolecules that could cause the adverse outcome resulting in liver damage. The combination of drugs or compounds (*i.e.,* lipids) carrying fragments susceptible to free radical oxidation and fragments causing the inhibition of signaling pathways meant to alleviate or prevent oxidative stress can all lead to liver damage. These MIEs and their adverse outcome(s) are described in the following paragraphs and are illustrated in Figure 4.

The assay AID 686978 identifies inhibitors of human TDP1. TDP1 is an enzyme that repairs single-stranded DNA breaks covalently linked to topoisomerase I. It is known that mutations in TDP1 impair the ability of a cell to repair DNA damaged by oxidation or drugs (Ben Hassine and Arcangioli 2009). When DNA is damaged and TDP1 is inhibited, topoisomerase I stays covalently linked to the DNA during replication and the cell dies (Pouliot et al. 1999). Since the ARE pathway contains a considerable number of detoxifying genes, it acts as the first line of defense to prevent DNA damage from oxidation or drugs (Kwak et al. 2003).

For AID 743067, active compounds in this assay act as TR antagonist and can disrupt metabolic homeostasis by inhibiting the binding of the thyroid hormone (Jameson and Weetman 2012). The liver plays a major role in thyroid hormone metabolism and liver damage is often associated with thyroid diseases (Huang and Liaw 1995). Furthermore, the liver metabolizes lipids and thyroid hormones regulate hepatic lipid homeostasis (Malik and Hodgson 2002). Lipids autoxidize in the presence of molecular oxygen, a process known as lipid peroxidation (Porter et al. 1995), which forms free radicals and ROS. Normally the ARE will inactivate ROS

(Shukla et al. 2012). Failure to terminate ROS results in oxidative stress (Sies 1997), especially when a TR antagonist has disrupted liver lipid metabolism.

The assay AID 743140 identifies PPARg agonists that activate the PPAR response elements and in this specific case it regulates adipogenesis (Tontonoz et al. 1994). Adipose tissue, especially visceral adipose tissue, releases fatty acids directly into the liver *via* the hepatic portal vein (Lafontan and Girard 2008). Fatty acids are susceptible to lipid peroxidation. Disrupting PPARg and adipogenesis could put the liver at risk for oxidative stress when fatty acids are in excess.

The AOP concept was presented as a logical sequence of biological responses that is useful for understanding complex toxicity phenomena (Allen et al. 2014; Ankley et al. 2010). Based on the AOP concept, Allen et al. discussed a unified MIE definition for the AOP framework for risk assessment purposes (Allen et al. 2014). This kind of research classifies compounds by mode of action using *in vitro* methods. Therefore, the chemical *in vitro-in vivo* relationships identified in this study can also be integrated into the AOP framework of liver damage. Potential toxicophore A is an electrophilic fragment highly susceptible to free radical oxidation, due to its allylic hydrogen (Porter et al. 1995). It represents a key chemical property of potential toxicants in an AOP framework. For example, oxyphenbutazone (CID 4641) is known for causing liver damage (Gaisford 1962). It contains potential toxicophore A and is active in AIDs 686978 and 743202 as a TDP1 inhibitor and ARE agonist, respectively. The bioassay results can be viewed as the macro-molecular interactions and the RA value can be considered as a specific cellular response pathway perturbation score (*i.e.,* ARE signaling pathway perturbation and TDP1 inhibition) of AOP for this compound. The molecular mechanism by which oxyphenbutazone causes liver damage is still not clear (Gaisford 1962; Tai

2012). However, it is well established that it is a lipid soluble drug metabolized by liver microsomal enzymes and requires molecular oxygen to metabolize (Davies and Thorgeirsson 1971). Similarly, potential toxicophore B is known as N-methylformamide, a well-known liver toxicant susceptible to free radical oxidation by C-H abstraction from alkyl group(s) adjacent to the nitrogen atom (Borduas et al. 2015). This reaction produces methyl isocyanate, which is highly toxic (Varma 1987). For example, 5-fluorouracil (CID 3385) contains toxicophore B. 5-fluorouracil was shown to be active in both AIDs 686978 and 743067, TDP1 inhibitor and TR antagonist, respectively. If administered orally, 5-fluorouracil is metabolically degraded predominantly in the liver by dihydropyrimidine dehydrogenase (DPD) (Omura 2003). Patients that lack DPD are highly likely to experience liver damage (Chabner et al. 2011). In our current study, it is noticeable that the four major components of an AOP (as defined by Ankley et al. 2010) are included: chemical properties of toxicants, macro-molecular interactions, cellular responses, and organ responses. Our future study will focus on the AOP framework of liver damage by differentiating the hepatotoxicity mechanisms of liver damage (*e.g.,* acute hepatic failure, cytolytic hepatitis, hepatic necrosis) (Zhu and Kruhlak 2014).

Our findings suggest that the four assays (686978, 743067, 743140, and 743202) could be used to screen for compounds that cause oxidative stress and induce liver damage. When specific chemical features (*e.g.,* potential toxicophores A and B) are present, the active responses obtained from these bioassays suggest potential hepatotoxicity. Although the four assays have covered several important mechanisms of oxidative stress, the negative results from all four assays would not be sufficient to indicate that a chemical is not hepatotoxic. Future work on this project includes the validation of these assays for their predictivity of liver damage, which will be used to optimize predictive liver toxicity models.

Conclusions

We developed a workflow that identified potential assays from a public big data source for the evaluation of liver damage caused by oxidative stress. Although using four assays will not be enough to cover all the relevant toxicity mechanisms of liver damage, this work clearly indicates the benefits of searching for useful toxicity data in the public big data domain for the compounds of interest. The increase in false positives in the IVIVC analysis indicates that the bioassay data is still needed for the compounds out of AD (*e.g.,* drug-like compounds). This issue could be resolved by rational design of the HTS chemical library that covers all the chemical space. New compounds containing the potential toxicophores can be tested using these four assays to assess the potential liver damage caused by oxidative stress prior to animal testing.

The workflow developed in this study can be easily adapted to study the relationship between any bioassay and other *in vivo* exposure data to evaluate complex *in vitro-in vivo* relationships and reveal toxicity mechanisms. Future directions of *in silico* modeling of animal toxicity induced by drugs and oxidative stress could include pharmacology studies.

References

Adler S, Basketter D, Creton S, Pelkonen O, van Benthem J, Zuang V, et al. 2011. Alternative (non-animal) methods for cosmetics testing: current status and future prospects-2010. Arch Toxicol 85:367–485.

Allen TEH, Goodman JM, Gutsell S, Russell PJ. 2014. Defining Molecular Initiating Events in the Adverse Outcome Pathway Framework for Risk Assessment. Chem Res Toxicol 27:2100–2112.

Ankley GT, Bennett RS, Erickson RJ, Hoff DJ, Hornung MW, Johnson RD, et al. 2010. Adverse outcome pathways: a conceptual framework to support ecotoxicology research and risk assessment. Environ Toxicol Chem 29:730–41.

Attene-Ramos MS, Miller N, Huang R, Michael S, Itkin M, Kavlock RJ, et al. 2013. The Tox21 robotic platform for the assessment of environmental chemicals - from vision to reality. Drug Discov Today 18:716–23.

Baell JB, Holloway GA. 2010. New substructure filters for removal of pan assay interference compounds (PAINS) from screening libraries and for their exclusion in bioassays. J Med Chem 53:2719–2740.

Ben Hassine S, Arcangioli B. 2009. Tdp1 protects against oxidative DNA damage in non-dividing fission yeast. EMBO J 28:632–40.

Borduas N, da Silva G, Murphy JG, Abbatt JPD. 2015. Experimental and Theoretical Understanding of the Gas Phase Oxidation of Atmospheric Amides with OH Radicals: Kinetics, Products, and Mechanisms. J Phys Chem A 119:4298–4308.

Breiman L. 2001. Random forests. Mach Learn 45:5–32.

Chabner BA, Bertino J, Cleary J, Ortiz T, Lane A, Supko JG, et al. 2011. Chapter 61. Cytotoxic Agents. In *Goodman & Gilman's The Pharmacological Basis of Therapeutics, 12e* (L.L. Brunton, B.A. Chabner, and B.C. Knollmanneds.), The McGraw-Hill Companies, New York, NY.

Chen M, Borlak J, Tong W. 2014. Predicting idiosyncratic drug-induced liver injury - some recent advances. Expert Rev Gastroenterol Hepatol 8:721–3.

Collins FS, Gray GM, Bucher JR. 2008. Toxicology. Transforming environmental health protection. Science 319:906–7.

Committee on Toxicity Testing and Assessment of Environmental Agents NRC. 2007. *Toxicity Testing in the 21st Century: A Vision and a Strategy*. The National Academies Press, Washington, DC.

Daniel WW. 2009. *Biostatistics: A Foundation for Analysis in the Health Sciences*. 9th ed. Wiley, Hoboken, NJ.

Davies DS, Thorgeirsson SS. 1971. Mechanism of hepatic drug oxidation and its relationship to individual differences in rates of oxidation in man. Ann N Y Acad Sci 179:411–20.

Dix DJ, Houck KA, Martin MT, Richard AM, Setzer RW, Kavlock RJ. 2007. The ToxCast program for prioritizing toxicity testing of environmental chemicals. Toxicol Sci 95:5–12.

Eaton DL, Gilbert SG. 2010. Chapter 2. Principles of Toxicology. In *Casarett & Doull's Essentials of Toxicology, 2e* (C.D. Klaassen and J.B. Watkinseds.), The McGraw-Hill Companies, New York, NY.

Ekins S. 2014. Progress in computational toxicology. J Pharmacol Toxicol Methods 69:115–40.

Fielden MR, Brennan R, Gollub J. 2007. A gene expression biomarker provides early prediction and mechanistic assessment of hepatic tumor induction by nongenotoxic chemicals. Toxicol Sci 99:90–100.

Gaisford W. 1962. Fatality after oxyphenbutazone in Still's disease. Br Med J 2: 1517.

Golbraikh A, Shen M, Xiao Z, Xiao Y-D, Lee K-H, Tropsha A. 2003. Rational selection of training and test sets for the development of validated QSAR models. J Comput Aided Mol Des 17:241–53.

Hartung T. 2009. Toxicology for the twenty-first century. Nature 460:208–12.

Huang MJ, Liaw YF. 1995. Clinical associations between thyroid and liver diseases. J Gastroenterol Hepatol 10: 344–50.

Huang R, Southall N, Wang Y, Yasgar A, Shinn P, Jadhav A, et al. 2011. The NCGC Pharmaceutical Collection: A Comprehensive Resource of Clinically Approved Drugs Enabling Repurposing and Chemical Genomics. Sci Transl Med 3:80ps16–80ps16.

Hybertson BM, Gao B, Bose SK, McCord JM. 2011. Oxidative stress in health and disease: the therapeutic potential of Nrf2 activation. Mol Aspects Med 32:234–46.

Inglese J, Auld DS, Jadhav A, Johnson RL, Simeonov A, Yasgar A, et al. 2006. Quantitative high-throughput screening: a titration-based approach that efficiently identifies biological activities in large chemical libraries. Proc Natl Acad Sci U S A 103:11473–8.

Itoh K, Chiba T, Takahashi S, Ishii T, Igarashi K, Katoh Y, et al. 1997. An Nrf2/small Maf heterodimer mediates the induction of phase II detoxifying enzyme genes through antioxidant response elements. Biochem Biophys Res Commun 236: 313–22.

Jameson JL, Weetman AP. 2012. Chapter 341. Disorders of the Thyroid Gland. In *Harrison's Principles of Internal Medicine, 18e* (D.L. Longo, A.S. Fauci, D.L. Kasper, S.L. Hauser, J.L. Jameson, and J. Loscalzoeds.), The McGraw-Hill Companies, New York, NY.

Kensler TW, Wakabayashi N, Biswal S. 2007. Cell survival responses to environmental stresses via the Keap1-Nrf2-ARE pathway. Annu Rev Pharmacol Toxicol 47:89–116.

Kim MT, Sedykh A, Chakravarti SK, Saiakhov RD, Zhu H. 2014. Critical evaluation of human oral bioavailability for pharmaceutical drugs by using various cheminformatics approaches. Pharm Res 31:1002–14.

Königer C, Wingert I, Marsmann M, Rösler C, Beck J, Nassal M. 2014. Involvement of the host DNA-repair enzyme TDP2 in formation of the covalently closed circular DNA persistence reservoir of hepatitis B viruses. Proc Natl Acad Sci U S A 111:E4244–53.

Kwak M-K, Wakabayashi N, Itoh K, Motohashi H, Yamamoto M, Kensler TW. 2003. Modulation of gene expression by cancer chemopreventive dithiolethiones through the Keap1-Nrf2 pathway. Identification of novel gene clusters for cell survival. J Biol Chem 278:8135–45.

Lafontan M, Girard J. 2008. Impact of visceral adipose tissue on liver metabolism. Part I: heterogeneity of adipose tissue and functional properties of visceral adipose tissue. Diabetes Metab 34:317–27.

Low Y, Uehara T, Minowa Y, Yamada H, Ohno Y, Urushidani T, et al. 2011. Predicting drug-induced hepatotoxicity using QSAR and toxicogenomics approaches. Chem Res Toxicol 24:1251–62.

Ma Q. 2013. Role of nrf2 in oxidative stress and toxicity. Annu Rev Pharmacol Toxicol 53:401–26.

MacDonald JS, Robertson RT. 2009. Toxicity testing in the 21st century: a view from the pharmaceutical industry. Toxicol Sci 110:40–6.

Malik R, Hodgson H. 2002. The relationship between the thyroid gland and the liver. QJM 95: 559–69.

Mantena SK, King AL, Andringa KK, Eccleston HB, Bailey SM. 2008. Mitochondrial dysfunction and oxidative stress in the pathogenesis of alcohol- and obesity-induced fatty liver diseases. Free Radic Biol Med 44:1259–72.

Moeller TA. 2010. From in vitro to in vivo: Preclinical assays to assess drug-induced liver injury demonstrate in vitro-in vivo correlation. Drug Discov Dev 13: 10–13.

O'Brien PJ, Irwin W, Diaz D, Howard-Cofield E, Krejsa CM, Slaughter MR, et al. 2006. High concordance of drug-induced human hepatotoxicity with in vitro cytotoxicity measured in a novel cell-based model using high content screening. Arch Toxicol 80:580–604.

Omura K. 2003. Clinical implications of dihydropyrimidine dehydrogenase (DPD) activity in 5-FU-based chemotherapy: Mutations in the DPD gene, and DPD inhibitory fluoropyrimidines. Int J Clin Oncol 8:132–138.

Porter NA, Caldwell SE, Mills KA. 1995. Mechanisms of free radical oxidation of unsaturated lipids. Lipids 30:277–290.

Pouliot JJ, Yao KC, Robertson CA, Nash HA. 1999. Yeast gene for a Tyr-DNA phosphodiesterase that repairs topoisomerase I complexes. Science 286:552–555.

Sedykh A, Zhu H, Tang H, Zhang L, Richard A, Rusyn I, et al. 2011. Use of in vitro HTS-derived concentration-response data as biological descriptors improves the accuracy of QSAR models of in vivo toxicity. Environ Health Perspect 119:364–70.

Shuhendler AJ, Pu K, Cui L, Uetrecht JP, Rao J. 2014. Real-time imaging of oxidative and nitrosative stress in the liver of live animals for drug-toxicity testing. Nat Biotechnol 32:373–80.

Shukla SJ, Huang R, Simmons SO, Tice RR, Witt KL, Vanleer D, et al. 2012. Profiling environmental chemicals for activity in the antioxidant response element signaling pathway using a high throughput screening approach. Environ Health Perspect 120:1150–6.

Sies H. 1997. Oxidative stress: oxidants and antioxidants. Exp Physiol 82:291–295.

Simmons SO, Fan C-Y, Yeoman K, Wakefield J, Ramabhadran R. 2011. NRF2 Oxidative Stress Induced by Heavy Metals is Cell Type Dependent. Curr Chem Genomics 5:1–12.

Tai WW. 2012. Chapter 117. Nonsteroidal Anti-Inflammatory Drugs. In *Poisoning & Drug Overdose, 6e.* (K.R. Olsoned.), The McGraw-Hill Companies, New York, NY.

Tice RR, Austin CP, Kavlock RJ, Bucher JR. 2013. Improving the human hazard characterization of chemicals: a Tox21 update. Environ Health Perspect 121:756–65.

Tontonoz P, Graves RA, Budavari AI, Erdjument-Bromage H, Lui M, Hu E, et al. 1994. Adipocyte-specific transcription factor ARF6 is a heterodimeric complex of two nuclear hormone receptors, PPAR7 and RXRa. Nucleic Acids Res 22:5628–5634.

Tropsha A, Golbraikh A. 2007. Predictive QSAR Modeling Workflow, Model Applicability Domains, and Virtual Screening. Curr Pharm Des 13:3494–3504.

Vapnik VN. 2000. *The Nature of Statistical Learning Theory*. Springer New York, New York, NY.

Varma DR. 1987. Epidemiological and experimental studies on the effects of methyl isocyanate on the course of pregnancy. Environ Health Perspect 72:153–157.

Venugopal R, Jaiswal AK. 1998. Nrf2 and Nrf1 in association with Jun proteins regulate antioxidant response element-mediated expression and coordinated induction of genes encoding detoxifying enzymes. Oncogene 17:3145–56.

Wang Y, Suzek T, Zhang J, Wang J, He S, Cheng T, et al. 2014. PubChem BioAssay: 2014 update. Nucleic Acids Res 42:D1075–82.

Willett P, Barnard JJM, Downs GMG. 1998. Chemical Similarity Searching. J Chem Inf Model 38:983–996.

Zhang D, Hannink M. 2003. Distinct cysteine residues in Keap1 are required for Keap1-dependent ubiquitination of Nrf2 and for stabilization of Nrf2 by chemopreventive agents and oxidative. Mol Cell Biol 23:8137–8151.

Zhang J, Hsieh J-H, Zhu H. 2014. Profiling animal toxicants by automatically mining public bioassay data: a big data approach for computational toxicology. PLoS One 9:e99863.

Zheng W, Tropsha A. 2000. Novel variable selection quantitative structure--property relationship approach based on the k-nearest-neighbor principle. J Chem Inf Comput Sci 40:185–94.

Zhu H, Kim M, Zhang L, Sedykh A. 2014a. CHAPTER 5. Computers Instead of Cells: Computational Modeling of Chemical Toxicity. In *Reducing, Refining and Replacing the Use of Animals in Toxicity Testing* (D. Allen and M.D. Waterseds.), pp. 163–182, The Royal Society of Chemistry, Cambridge.

Zhu H, Zhang J, Kim MT, Boison A, Sedykh A, Moran K. 2014b. Big data in chemical toxicity research: the use of high-throughput screening assays to identify potential toxicants. Chem Res Toxicol 27:1643–51.

Zhu X, Kruhlak NL. 2014. Construction and analysis of a human hepatotoxicity database suitable for QSAR modeling using post-market safety data. Toxicology 321:62–72.

Table 1. Comprehensive toxicity databases compiled from public sources

Names	Types	Description	Number of compounds
Tox21 phase I (NTP and EPA) ARE-*bla* (https://pubchem.ncbi.nlm.nih.gov/assay/assay.cgi?aid=651741, accessed August 29, 2015)	*In vitro*	Compounds characterized in traditional toxicology tests and/or known to be harmful to humans and the environment	2,617
Tox21 phase II 10K ARE-*bla* (http://www.epa.gov/ncct/dsstox/sdf_tox21s.html, accessed October 2, 2012)	*In vitro*	Diverse compounds (pesticides, industrial, food-use, drugs, etc.) with chemical features that are of interest to toxicologists	8,311
FDA liver damage (Zhu and Kruhlak 2014)	*In vivo*	Drugs known to cause liver damage (*e.g.,* necrosis, lesions, traumatic liver injury)	1,314
PubChem Bioassay (http://www.ncbi.nlm.nih.gov/pcassay/, accessed February 27, 2014)	*In vitro & in vivo*	Compounds that have been validated and screened in different bioassays	48M+

Table 2. Definition of compound activity categories from concentration-response curves and the CurveP algorithm for the qHTS ARE-*bla* datasets

Category	Activity	CurveP	Maximum response	Number of responses > 20 units
Active[a]	1	> 0	≥ 20	> 1
Potential active[b]	0.75	> 0	≥ 20	= 1
Inconclusive[c]	0.25	= 0	< 20	= 0
Inactive[d]	0	= 0	< 10	= 0

[a]Strong ARE-*bla* activation signals observed; [b]Weak ARE-*bla* activation signal observed; [c]Inconsistent ARE-*bla* activation signal(s) observed; [d]Negligible or no ARE-*bla* activation signals observed.

Table 3. qHTS ARE-*bla* combinatorial QSAR consensus model performance in 5-fold cross validation and against external validation sets, with and without Applicability Domain (AD), Consensus Prediction Thresholds (CPT) 1-2

Statistics		5-fold cross validation (80/20% split)	Validation set I	Validation set I + AD	Validation set II	Validation II + AD
n (active/inactive)		750/800	175/973	132/757	128/3,456	59/2,566
CPT-1[a]	Sens[c] (%)	76	76	73	83	80
	Spec[d] (%)	71	83	85	72	78
	CCR[e] (%)	74	80	79	77	79
	Coverage[f] (%)	100	100	77	100	73
CPT-2[b]	Sens. (%)	75	68	62	93	90
	Spec. (%)	92	99	99	92	95
	CCR (%)	84	84	80	92	93
	Coverage (%)	35	40	34	45	37

[a]CPT-1:QSAR prediction ≥0.5 as actives and QSAR prediction < 0.5 as inactives; [b]CPT-2: QSAR prediction ≥ 0.8 as actives and QSAR prediction ≤ 0.3 as inactives; [c]Sens, sensitivity - percentage of active or toxic compounds predicted correctly; [d]Spec, specificity - percentage of inactive or non-toxic compounds predicted correctly; [e]CCR, correct classification rate; [f]Coverage - fraction of compounds that are within the applicability domain.

Figure Legends

Figure 1. The workflow for profiling liver toxicants consists of three major stages: (1) automated biological response profiling, (2) QSAR modeling of qHTS ARE-*bla* activation, (3) chemical IVIVC evaluation. In the columns [Liver Damage, 1, 2, 3, "…", *n,* ARE-*bla*], actives are red color and "1;" inactives are blue and "0;" and inconclusive or untested are white and empty.

Figure 2. Chemical space plot of (A) the modeling set (actives = red, inactives = purple) vs. its left out compounds (yellow) and (B) the modeling set vs the FDA liver damage compounds (green) using the top three principal components generated using 186 MOE 2-D descriptors.

Figure 3. The IVIVC between selected assays and liver damage was evaluated by classifying responses as true positive (TP), true negative (TN), false positive (FP), or false negative (FN) for a X^2 ($\alpha = 0.05$) or CCR test. (A) The biological response profile (red = active or toxic, blue = inactive or non-toxic, yellow = inconclusive or untested) of liver damage compounds represented in the heat map using the top four assays (AIDs 686978, 743067, 743140, and 743202). Individual assays show weak IVIVC, but the combined responses of the assays using threshold RA > 0.25 as active resulted in a statistically significant IVIVC (X^2 p-value = 2.92×10^{-4}). (B) The IVIVC between experimental qHTS ARE-bla activation and liver damage and the QSAR predictions for each liver damage compound, for subsets of overlapping compounds with potential toxicophores A (left) and B (right).

Figure 4. The potential liver toxicity mechanism of the compounds, like oxyphenbutazone (CID 4641) and 5-fluorouracil (CID 3385), that contain either of the proposed toxicophores A or B can generate reactive oxygen species. These types of stimuli activate the Antioxidant Response Element signaling pathway (ARE) (AID 743202) and peroxisome proliferator-activated receptor

gamma signaling pathway (PPARg) (AID 743140), inhibit human tyrosyl-DNA phosphodiesterase 1 signaling pathway (TDP1) (686978), or disrupt the thyroid receptor signaling pathway (TR) (AID 743067).

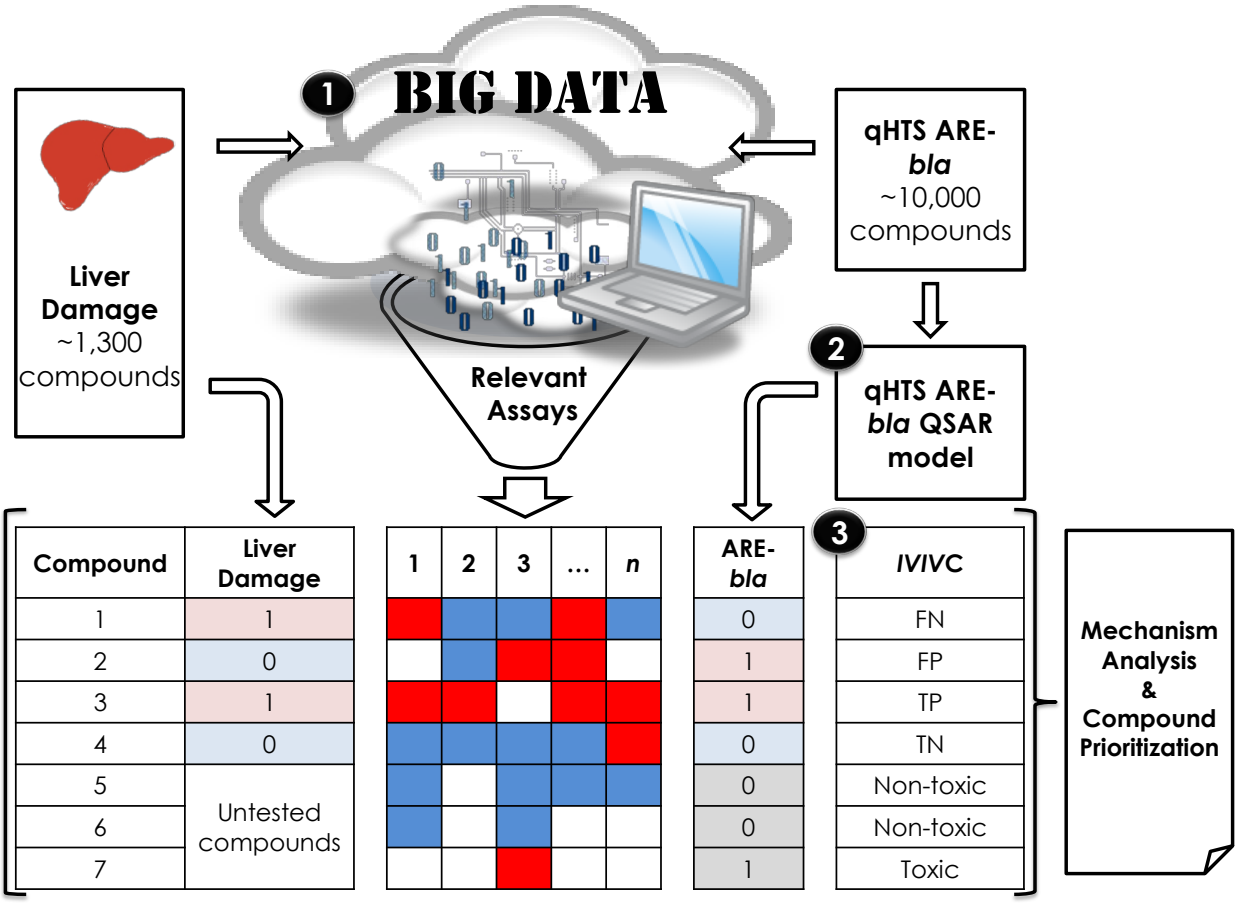
1
BIG DATA
Liver Damage ~1,300 compounds
qHTS ARE-bla ~10,000 compounds
Relevant Assays
2
qHTS ARE-bla QSAR model
3
Compound
Liver Damage
1
0
1
0
Untested compounds
1
2
3
…
n
ARE-bla
0
1
1
0
0
0
1
IVIVC
FN
FP
TP
TN
Non-toxic
Non-toxic
Toxic
Mechanism Analysis & Compound Prioritization

Figure 1.

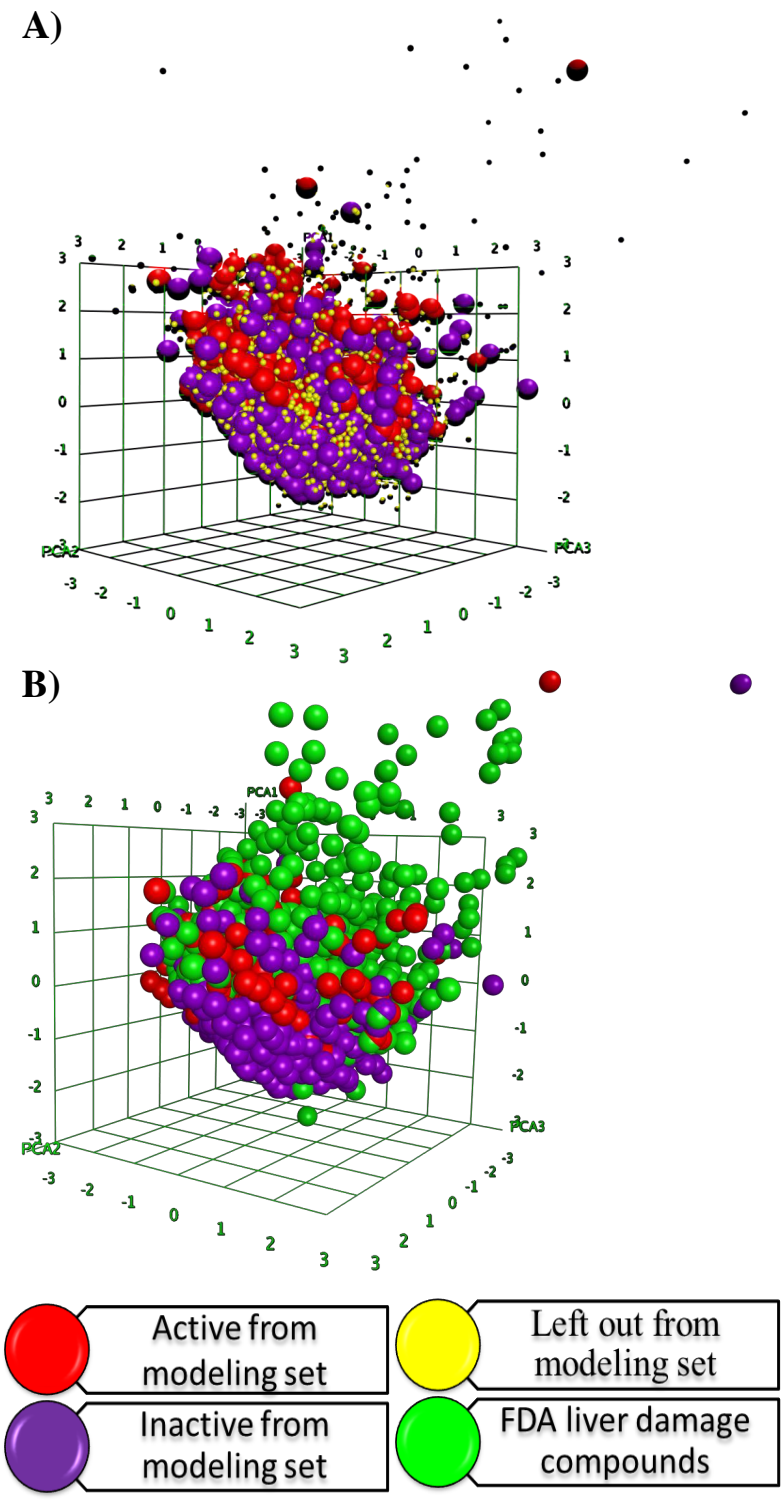
A)
B)
PCA1
PCA2
PCA3
Active from modeling set
Left out from modeling set
Inactive from modeling set
FDA liver damage compounds

Figure 2.

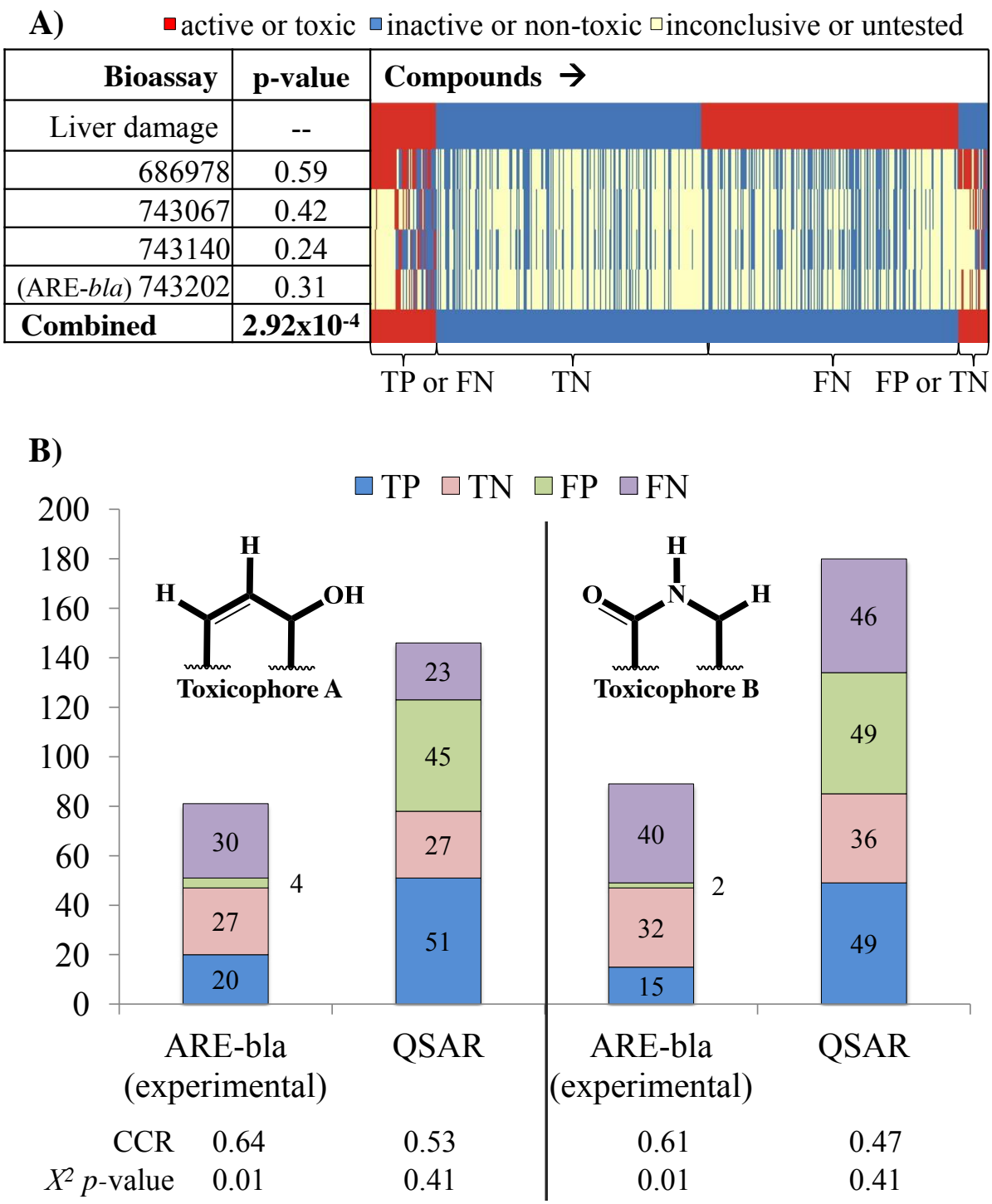
A)
active or toxic
inactive or non-toxic
inconclusive or untested
Bioassay
p-value
Compounds →
Liver damage
--
686978
0.59
743067
0.42
743140
0.24
(ARE-bla) 743202
0.31
Combined
2.92x10-4
TP or FN
TN
FN
FP or TN
B)
TP
TN
FP
FN
200
180
160
140
120
100
80
60
40
20
0
Toxicophore A
Toxicophore B
30
27
20
4
23
45
27
51
40
32
15
2
46
49
36
49
ARE-bla (experimental)
QSAR
ARE-bla (experimental)
QSAR
CCR
0.64
0.53
0.61
0.47
X2 p-value
0.01
0.41
0.01
0.41

Figure 3.

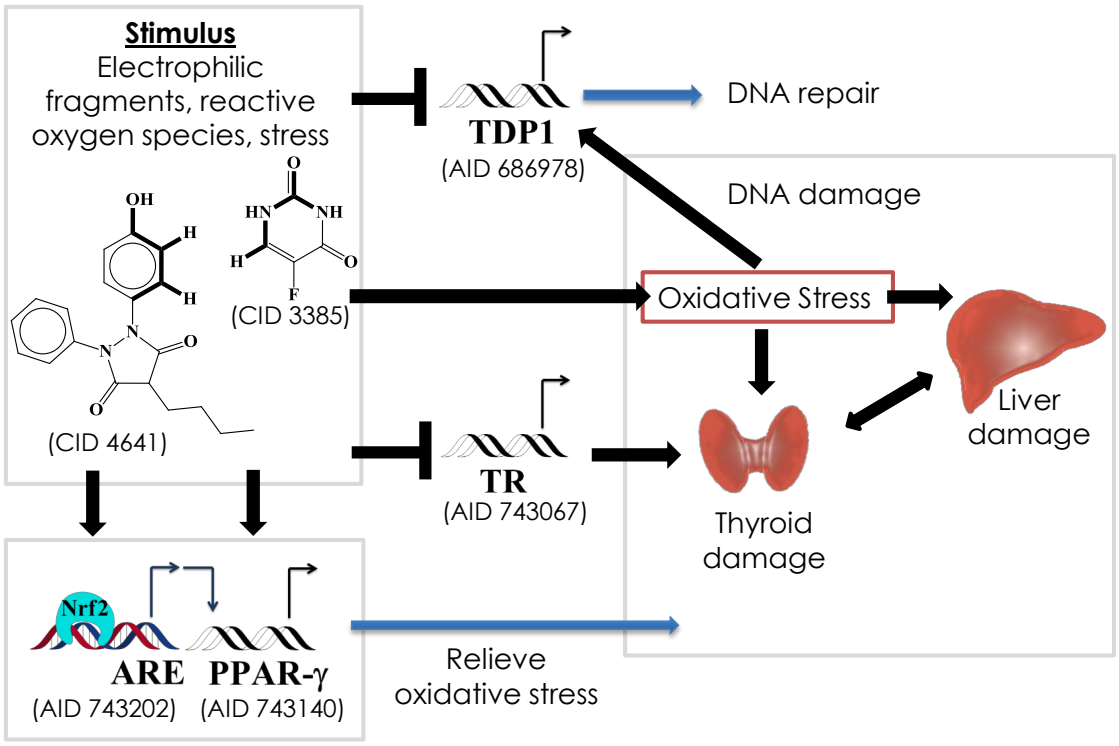
Stimulus
Electrophilic fragments, reactive oxygen species, stress
(CID 4641)
(CID 3385)
TDP1
(AID 686978)
DNA repair
DNA damage
Oxidative Stress
Liver damage
TR
(AID 743067)
Thyroid damage
Nrf2
ARE
(AID 743202)
PPAR-γ
(AID 743140)
Relieve oxidative stress

Figure 4.